**IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF MISSISSIPPI
NORTHERN DIVISION**

EMMERICH NEWSPAPERS, INCORPORATED	**PLAINTIFF**
VS.	**CIVIL ACTION NO. 3:23CV26 TSL-MTP**
PARTICLE MEDIA, INC. D/B/A NEWS BREAK and JOHN DOES 1-10	**DEFENDANTS**

CONSOLIDATED WITH

EMMERICH NEWSPAPERS, INCORPORATED	**PLAINTIFF**
VS.	**CIVIL ACTION NO. 3:23CV391 TSL-MTP**
PARTICLE MEDIA, INC. D/B/A NEWS BREAK	**DEFENDANTS**

PLAINTIFF'S MEMORANDUM IN SUPPORT OF ITS MOTION TO EXCLUDE THE REPORT AND TESTIMONY OF MARK GIANTURCO

The Report and Testimony of Plaintiff's expert Mark Gianturco should be excluded on grounds the reasoning behind his analysis is unsound and cannot properly be applied to the facts in issue in this case. *See Daubert v. Merrell Dow Pharmaceuticals, Inc.,* 509 U.S. 579 (1993).

Particle Media retained Mr. Gianturco to analyze Emmerich's paywalls and give an expert opinion as to whether they were effective in the ordinary course of their operation. Not surprisingly, he concluded that they were not. However, the methodology he employed cannot withstand scrutiny. He relied on insufficient facts and impermissibly flawed principles and methods which dictate the exclusion of his testimony.

Emmerich alleges that Particle violated 17 U.S.C. 1201(a)(1)(A), which states, "No person shall circumvent a technological measure that effectively controls access to a work protected under this title." To "circumvent a technological measure" means to "descramble a scrambled work, to decrypt an encrypted work, ***or otherwise to avoid, bypass, remove, deactivate, or impair a technological measure***, without the authority of the copyright owner" 17 U.S.C. 1201(a)(3)(a)(emphasis added). And a technological measure "effectively controls access to a work" if the measure, in the ordinary course of its operation, requires the application of information, or a process or a treatment, with the authority of the copyright owner, to gain access to the work. 17 U.S.C. 1201(a)(3)(b).

During the time frame in question a number of Emmerich websites (including, among others, the *Greenwood Commonwealth* and the *McComb Enterprise-Journal*) operated behind paywalls. A reader could only view full articles by paying a subscription fee. These paywalls were specifically intended to limit access to paying customers and motivate prospective readers to buy subscriptions. Ex. "A", Emmerich Affidavit ¶¶10-11; Ex. "B", Ryan Affidavit ¶¶6-11; Ex. "C" Kalich Affidavit ¶¶6-12.

Particle utilized a web crawler to systematically steal content from Emmerich's websites (a process known as "scraping"). Particle's web crawler would scrape the entire web page for Emmerich articles, remove everything except the headline, image and news story, and then store that content in Particle's database. *See* Ex. D, First Zhong Deposition Transcript, p. 16-17. Snippets of these articles appeared on the NewsBreak news feed and readers "clicking through" were led to either a full-text view of the stripped-down Emmerich article in "self-hosted" mode or a Framed View, in which the entire Emmerich article appeared in a frame or window in the Particle app.

The Greenwood Commonwealth and the McComb Enterprise Journal used a content management system called TownNews (now known as Blox). During the relevant time period Emmerich's other websites utilized an open-source system called Drupal. TownNews is one of the largest content management systems in the newspapers industry, with over 2,000 news sites representing over one quarter of all online news content providers**.** TownNews and Drupal perform the same function but in different ways**.**[1] *See* Ex. E, Second Wyatt Emmerich Affidavit, ¶ 3.

The TownNews paywall generated a pop-up block when a reader tried to click-through. This block required the reader to subscribe before viewing the article, and assigned a unique user ID and password to the reader. This paywall was effective in the ordinary course of business and Emmerich generated millions of dollars in subscription revenue as a result of this paywall. *See* Ex. E, Second Wyatt Emmerich Affidavit, ¶ 4.

The TownNews paywall consisted of approximately 300 lines of HTML code inserted automatically by TownNews into every online Emmerich article. This HTML code appeared in every web page containing an Emmerich article. Emmerich went to great lengths to optimize the performance of its websites to ensure the paywalls worked properly and the *"subscriber-only, encrypted-content"* block appeared almost instantaneously. The Particle parser plucked Emmerich's headline, image and content but deleted the 300 lines of HTML code that created Emmerich's paywall. This was deletion and impairment of a copyright protection system and meets the statutory definition of circumvention in 1201. *See* Ex. E, Second Wyatt Emmerich Affidavit, ¶ 5.

---

[1] Particle claims that because its web crawler was blocked at Emmerich websites utilizing Drupal, this means Greenwood and McComb must *not* have been behind effective paywalls. However, as noted below, the TownNews paywalls utilized in Greenwood and McComb effectively controlled access to Emmerich's website in the ordinary course of their operation. They simply were not designed to prevent hacking by a sophisticated attacker like Particle.

An example of this HTML code inserted by TownNews to create a paywall in every article appears below:

```
19  { NodeList.prototype.forEach = Array.prototype.forEach; } /** IE11
20  polyfill */
         document.querySelectorAll(".subscriber-only.encrypted-
21  content").forEach(function(el){
             el.innerHTML = tncms.unscramble(el.textContent);
22           el.classList.remove("encrypted-content");
         });
23       document.querySelectorAll(".subscriber-only,.subscriber-
24  hide").forEach(function(el){ el.style.display = "" });
         __tnt.subscription.api.decrypt = function(){ return false; }
25     };
```

"[P]assword protection . . . and validation keys are technological measures within the meaning of the DMCA." *LivePerson, Inc. v. 24/7 Customer, Inc.*, 83 F. Supp. 3d 501, 510, 2015 U.S. Dist. LEXIS 3688, *12, 2015 WL 170348. Furthermore, "[T]he legal standard does not require airtight protection." *Yout, LLC v. Recording Indus. Ass'n of Am., Inc.*, 633 F. Supp. 3d 650, 670, 2022 U.S. Dist. LEXIS 178462, *38-39. "[A] technological measure need not establish an impenetrable barrier around a protected work to be 'effective' as a matter of law." Id. at *24. An ordinary consumer would have been blocked by Emmerich's paywalls at these websites, which is all §1201(a)(3)(B) requires.

During discovery Particle described a three-step process by which it acquires content for its app/website. *See* Ex. D, First Zhong Depo. pp. 15-20. Step one is "scraping," i.e., using a web crawler to systemically seek out news sites and scrape the entire web page for articles. Each article is surrounded by vast amounts of HTML code including the TownNews paywall code -- a typical Emmerich news article will contain up to *150 pages* of HTML code. *See* Ex. E, Second Wyatt Emmerich Affidavit, ¶ 6. It is undisputed that Particle's web crawler scraped or consumed the entire web page, including the TownNews paywall code, every time an Emmerich article was scraped:

Q.   Okay. When the web crawler identifies a news article, does it consume the

> Jentire article or just portions of it?
>
> A.   **The input to the parser is a whole web page**, and we try to extract headline, image, and content.
>
> Q.   I want to make sure I understand the first part of your response.· Did you say that you -- you -- **the web crawler consumes the entire web page?**
>
> A.   **Yes**.
>
> Q.   Does the web crawler consume advertisements which are located within the body of the article?
>
> A.   **The input is a whole page.**

See Ex. D, First Zhong Depo., pp. 24-25 (emphasis added). By definition this means Particle's web crawler consumed the HTML code inserted by TownNews to create the paywall for Emmerich.

It is virtually impossible for website operators like Emmerich to prevent web crawlers from scraping their content. Google and other search engines do this constantly in order to analyze content and direct internet traffic to appropriate sites. Scraping is not a problem as long as the party doing the scraping (1) does not republish the content on their own platforms without permission and (2) does not make paywall-protected content available to non-subscribers. *See* Ex. E, Second Wyatt Emmerich Affidavit, ¶ 7. Particle did both of these things.

It is undisputed that Particle programmed its web crawler to only identify and "parse" the headline, image and content for each article, which they then saved on their server. Zhong explained the process:

> A.   Okay. Yeah. Yeah, this is the diagram. I will use the top part to answer this question. So yeah, there are multiple components in the system. Here we have crawler and the parser and the database and the APP. So for the crawler to work, we take the input, the RSS, and the web pages, and we download them. That's what the crawler does. Then the results go into the parser. The parser will [parse] the content, things like title, image, and content. It will also generate a few useful information like the category of the page. Then it will be saved -- in the database.

Ex. D, First Zhong Depo. pp. 16-17. Particle admits that all of the other HTML code, including the TownNews paywall code, was discarded by its parser:

> Q.   **What happens to the rest of the HTML code that does not fall into the category of headline, image, or content?**
>
> . . . .
>
> A.   Yeah. The parsing process, like I just explained, is to get necessary component to do the analysis and the moderation, yeah. We do not need other component in the parsing stage. We only -- because later we redirect the traffic to the original website, so we do not save other information. The parser component only tries to extract title, content, image . . .
>
> Q.   **And is it accurate to say that the rest of the code is discarded?**
>
> A.   What we do is to extract what we need in the parser and to state those companies. I think that's what the company is doing. Yeah. That is the right way of referring what this component does.
>
> Q.   And so you don't save the rest of the HTML code?
>
> A.   **We do not save the rest of the HTML.**

See Ex. F, Second Zhong Depo., pp. 15-16 (emphasis added). Particle intentionally disregarded the paywall HTML code. Zhong admitted that the parser could have been programmed to identify Emmerich's paywalls:

> Q.   And you're saying it would be impossible to program your parser to identify the paywall tags?
>
> MR. CARMODY: Object to the form.
>
> A.   No.

See Ex. G, Particle 30(b)(6) Depo, p. 64.

This is where Particle removed, deactivated and impaired Emmerich's paywall. The circumvention did *not* occur when Particle scraped Emmerich's articles, which included all of the *"subscriber-only, encrypted-content"* HTML code inserted to prevent non-subscribers from viewing the article. Instead, it occurred during the "parsing" process when Particle's system cherry-picked each article's headline, image and content and then left everything else on the cutting room floor, *including the TownNews paywall code.* Particle then saved the stolen content on its server where anyone using the NewsBreak app could read the article in its entirety without the inconvenience of paying. Based on Particle's data there were over 276,996 NewsBreak page views of Emmerich's paywall-protected content for the articles remaining in this lawsuit,

resulting in a direct loss in subscription revenue for Emmerich. *See* Ex. E, Second Wyatt Emmerich Affidavit, ¶ 8.

In order to refute Emmerich's proof of this obvious circumvention of its paywalls, Particle retained an expert, Mark Gianturco, who was able to download the HTML code for four Emmerich articles from a totally different website known as Wayback.org (a/k/a/ the "Wayback Machine").

This was a poor choice of comparison, since the manner in which the Wayback Machine gathers and disseminates material has explicitly been held to constitute copyright infringement. *See*, *Hachette Book Grp., Inc. v. Internet Archive*, 664 F. Supp. 3d 370, 2023 U.S. Dist. LEXIS 50749. "The defendant, Internet Archive ("IA"), is a non-profit organization dedicated to providing 'universal access to all knowledge' . . . One of IA's first projects was to document the history of the Internet by archiving every public webpage on the World Wide Web through IA's "Wayback Machine." *Hachette Book Grp., Inc. v. Internet Archive*, 664 F. Supp. 3d 370, 374, 2023 U.S. Dist. LEXIS 50749, *4.

However, Wayback did not limit its activities just to "public webpages," but also scanned and published *millions* of copyrighted works as well. "Relevant to this action, however, the Website also includes 3.6 million books protected by valid copyrights, including 33,000 of the Publishers' titles and all of the Works in Suit. . . . The Publishers did not authorize IA to create digital copies of the Works in Suit or to distribute those unauthorized ebook editions on IA's Website." 664 F. Supp. 3d at 375, 2023 U.S. Dist. LEXIS 50749, *7-8. In other words, the publishers who sued Wayback were in exactly the same posture as Emmerich. Like those publishers Emmerich never authorized Wayback to scrape and display its articles, but Wayback did so anyway. The *Hatchette* Court rejected Wayback's claim of fair use, stating, "What fair use

does not allow . . . is the mass reproduction and distribution of complete copyrighted works in a way that does not transform those works and that creates directly competing substitutes for the originals. Because that is what IA has done with respect to the Works in Suit, its defense of fair use fails as a matter of law." 664 F. Supp. 3d at 391, 2023 U.S. Dist. LEXIS 50749, *44-45, 2023 U.S.P.Q.2D (BNA) 367.

In other words, Wayback has been found liable for copyright infringement for doing precisely what Emmerich alleges Particle has done in this case. Again, an exceedingly poor choice for comparison but that is the route Particle has chosen to follow.

From Wayback, Gianturco obtained massive amounts of HTML code for each Emmerich article he analyzed, approximately 150 *pages* of HTML code per article. Of critical importance, he did not examine any of these articles on the Emmerich websites where they originated or see the manner in which the paywalls functioned at those locations. This is because Emmerich has migrated both websites to a different paywall protection system which performs the same function but in a different manner. As a result, it was impossible for Gianturco to make an apples-to-apples comparison. Nevertheless, he opined that since he was able to download and then locate the content of Emmerich articles buried within those 150 pages of HTML code, this proves Emmerich's paywalls were not effective in their ordinary course of operation. *See below*.

Mr. Gianturco's report and expert opinion are so riddled with deficiencies they cannot be admissible under *Daubert*. First, the 150 pages of HTML code retrieved by Mr. Gianturco is unintelligible to the ordinary reader. In fact, Mr. Gianturco candidly admitted that *he* did not know what much of the code represented or did**.** *See* Ex. I*,* Gianturco Deposition Transcript, p. 22. No ordinary reader will download 150 pages of HTML code in order to seek out five or six paragraphs of a news story they can read by paying a small subscription fee. *See* Ex. E, Second

Wyatt Emmerich Affidavit, ¶ 9.

Second, Mr. Gianturco acknowledged that he never examined any of the stolen articles on

Emmerich's original websites behind the paywalls as they worked in their original location.

```
 6      Q.  And did you ever have the opportunity
 7   to view any of these articles on native -- I'm
 8   sorry -- Emmerich's native website in their
 9   original form protected by a TownNews paywall?
10          MS. HOWELL:  Object to form.
11          THE WITNESS:  I don't think anyone
12   could ever view them protected by the paywall
13   from 2019 to 2021 because I don't believe it
14   effectively existed, but if your question is
15   simply did I see these in 2019 to 2021, no, I
16   did not.
17   BY MR. CARROLL:
18      Q.  Okay.  And so you've never been able
19   to examine it as it appeared at the time on
20   Emmerich's website, correct?
21          MS. HOWELL:  Object to form.
22          THE WITNESS:  Correct.
```

Ex. I, Gianturco Deposition Transcript, p. 30. And later he openly acknowledged that he could

not say for certain whether the paywalls would have worked effectively at Emmerich's original

websites:

```
 1      Q.  Again, that's the way it worked on the
 2   Wayback Machine, but you don't know for a fact
 3   that's the way it worked on the original
 4   Emmerich website, do you?
 5          MS. HOWELL:  Object to form.
 6      A.  Well, again, obviously I can't, you
 7   know, go back in time, so to the extent that you
 8   are making a representation based on what was
 9   happening at a point in time, I can't say that I
10   knew that.
```

Id. at p. 35. There is no way he can feasibly compare the effectiveness of Emmerich's paywalls

at their original site with their operation on an entirely different website which has been found

guilty of copyright infringement.

Third, Mr. Gianturco acknowledged that Wayback inserted thousands of lines of its own code along with Emmerich's and TownNews' code, and confirmed that he had no idea what that code did or how it interacted with Emmerich's native code. When confronted with the code inserted by Wayback he testified as follows:

```
 6  BY MR. CARROLL:
 7      Q.  Well, I'm sorry.  We just established
 8  that they inserted code.  We're looking at the
 9  middle of the page.  This is on page 45
10  beginning at line 21.
11  Https://web.archive.org/web and a whole bunch of
12  stuff in blue there, that's clearly inserted by
13  Wayback, wasn't it?
14          MS. HOWELL:  Object to form.
15          THE WITNESS:  No.
16  BY MR. CARROLL:
17      Q.  What's your answer?
18      A.  No.
19      Q.  Wayback did not insert the code
20  beginning on line 21 down through line 23 on
21  page 46?
22      A.  I'm disagreeing with your verb.
23      Q.  What -- what verb would you use?
24      A.  I would say that Wayback needs to have
25  a frame or a context within which to display the
```

Page 21

```
 1  content from the Emmerich website, and without
 2  that, what you would get when you visited the
 3  Wayback Machine was a copy of the Emmerich's
 4  website without any indication of the website
 5  that you were visiting, which is the Wayback
 6  Machine.  So that just as Google, when they show
 7  other results, will have information indicating
 8  you're on the Google site, you similarly have
 9  indication and information indicating that
10  you're on the Wayback site, and that is all this
11  is.
12       Q.  And how do they create that frame?
13       A.  I don't know.
14          MS. HOWELL:  Object to form.
15  BY MR. CARROLL:
16       Q.  By inserting code, correct?
17       A.  No.
18       Q.  All right.  Where did this code come
19  from that's on page 21 -- or line 21 through 23
20  on page 46 of your report?
21       A.  What do you mean by where did it come
22  from?
23       Q.  Well, did it come from Emmerich's
24  native website?
25       A.  No.
```

Id. at p. 21. He candidly acknowledged that Wayback had inserted its own code into the

Emmerich articles he reviewed, and that he could not distinguish between that code and

Emmerich's original code:

Page 18

1   A.   I think what I'm comfortable saying is
2 that certainly web.archive.org and those
3 references that you're referring to did not
4 exist on the newspaper website.
5   Q.   Perfect. Thank you. And can you
6 identify all of the HTML code that was
7 inserted -- not manipulated -- but inserted by
8 Wayback?
9   A.   I mean, on the fly, I can't. It's
10 certainly within my capabilities, but I am not
11 going to go through that sort of exercise on the
12 fly here.
13   Q.   Okay. But you didn't make any attempt
14 to do that in your report, did you?
15   MS. HOWELL: Object to form.
16   THE WITNESS: Purposely, I did not
17 manipulate anything I received and I presented
18 it as it was received.
19 BY MR. CARROLL:
20   Q.   I didn't ask if you manipulated it. I
21 asked you if you tried to distinguish between
22 the code that existed on Emmerich's original
23 website and the code that was inserted by
24 Wayback. Did you make any effort to make that
25 distinction?

Id. at 17-18. Gianturco candidly admitted he had no idea how Wayback generated the code which they inserted in the Emmerich articles or the manner in which that code interacted with Emmerich's original code, because he never gained access to Wayback's own server code which would explain the operation of its systems:

```
                                              Page 22
1       Q.  Where did it come from?  You're the
2   expert, not me.
3       A.  Yeah.  I --
4           MS. HOWELL:  Object to form.
5           THE WITNESS:  -- don't know the
6   technologies they used to generate that code.
7   When you say where does it come from, that is an
8   artifact that's generated by code executing on
9   the server.  I have no idea where it came from
10  and neither can anyone else who didn't see the
11  code that was written on there, their server to
12  generate this output.
13  BY MR. CARROLL:
14      Q.  Okay.  Are you aware of all of the
15  ways -- let's call this -- let's call this
16  "Wayback code" for lack of a better term.  Are
17  you aware of all of the ways that the Wayback
18  code interacts with Emmerich's native code?
19      A.  If the question is am I aware right
20  now, no.  Could I -- no.  I mean, I'm aware of
21  many of the ways.  I could not be aware of all
22  the ways.  Again, I would have to have their
23  server code to understand that.
```

Id. at p. 22.

Fourth, he admitted that he had no idea what steps Emmerich took to optimize the operation of its website to insure its paywalls worked effectively in their native format:

```
11      Q.  Are you familiar, Mark, with the ways
12  that Emmerich would optimize their websites to
13  move quickly in a way that Wayback would not?
14          MS. HOWELL:  Object to form.
15          THE WITNESS:  I mean, what I'm aware
16  of is what I was able to examine.  To the extent
17  that there are other things that are obfuscated
18  by their server architectures or, you know, web
19  font deployments or caching or whatever else, I
20  just wouldn't -- I wouldn't have access to it.
```

Id. at p. 32.

And fifth, he acknowledged that Emmerich's paywall *did* appear even on the Wayback

Machine, albeit sometimes in a delayed fashion:

```
 5      Q.  When we pulled up one of the articles
 6   that you cited, it says, "Food boxes to be given
 7   away Friday morning at the ag center."  And then
 8   virtually immediately, we get a statement that
 9   says, "Thank you for reading.  Please log in or
10   sign up for a new account and purchase a
11   subscription for as little as $1 to keep
12   reading."
13           Would you agree that is a paywall,
14   Mark?
15      A.  Well, I -- you definitely have me
16   confused because you told me you were going to
17   follow the process that I documented in my
18   report and then you did something completely
19   different.  And one thing that's important when
20   you're working with any website, sequence of
21   operations can give you very different results.
22   And I think one of the things that has been
23   clear is that people including the code on the
24   website relied on people taking certain
25   sequences of operations that they didn't always
```

Id. at p. 26. Gianturco was retained to opine on the operation of Emmerich's paywalls "in the ordinary course of their operation." But when asked whether an ordinary reader would be deterred by the paywall which appeared even in Wayback, he answered as follows:

```
 6      Q.  Okay.  And you're saying that as an
 7   expert in computers, but wouldn't an ordinary
 8   person look at that and say, Well, there's a
 9   paywall; I have to subscribe?
10      A.  I don't know what an ordinary person
11   would or wouldn't say.
```

Id. at 28. In other words, Gianturco frankly admitted he had no opinion on the central question he was retained to answer, i.e., whether an ordinary person would have been prevented by Emmerich's paywalls in the ordinary course of their operation from reading an Emmerich article.

"Rule 702 charges trial courts to act as 'gate-keepers,' making a 'preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of

whether that reasoning or methodology properly can be applied to the facts in issue.' In short, expert testimony is admissible only if it is both relevant and reliable. This gate-keeping obligation applies to all types of expert testimony, not just scientific testimony." *Pipitone v. Biomatrix, Inc.*, (5th Cir. 2002) 288 F.3d 239, 243-244.

Gianturco's failure to examine the operation of Emmerich's paywalls at their original location, unimpeded by thousands of lines of random HTML code inserted by Wayback in order to facilitate viewing the articles in question renders his report and testimony hopelessly unsound. For this reason the Court should exclude his report and testimony under the well-established rules of *Daubert*.

RESPECTFULLY SUBMITTED, this the 29th day of July, 2024.

BY: /s/ <u>Wilson H. Carroll</u>
Wilson H. Carroll (MSB#5894)

Wilson H. Carroll
WILSON CARROLL, PLLC
2506 Cherry Street
Vicksburg, Mississippi 39180
(601)953-6579
wilson@wilsoncarroll.com

CERTIFICATE OF SERVICE

I, Wilson H. Carroll, hereby certify that I have this day filed the foregoing document via the Court's automated filing system, which automatically forwarded copies to all counsel of record.

So certified, this the 29th day of July, 2024.

BY: /s/ <u>Wilson H. Carroll</u>
Wilson H. Carroll (MSB#5894)